GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:09:40 ; Search time 91.75 Seconds
                                            (without alignments)
                                            234.418 Million cell updates/sec

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Sequence: 1 MAINTSNLCSLLFLLSLFLL.............SPRSTKQQOPLVSILDFVGF 666
Perfect score: 3542
Title: US-09-331-631A-1

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description
1	1092	30.8	605	1	VCLA_GOSHI	P0799; gossypium h
2	1082.5	30.6	588	1	VCLB_GOSHI	P09801 gossypium h
3	905	25.6	573	1	GLB1_MAIZE	P15590 zea mays (m
4	856.5	24.2	639	1	GLCX_SOYBN	P11827 glycine max
5	849.5	24.0	626	1	AH12_ARAHY	P43238 arachis hyp
6	849	24.0	614	1	AH11_ARAHY	P43237 arachis hyp
7	840.5	23.7	605	1	GLCA_SOYBN	P13916 glycine max
8	821.5	23.2	571	1	CVCA_PEA	P13915 glycine max
9	808	22.8	524	1	SBP_SOYBN	Q04672 glycine max
10	745.5	21.0	459	1	VCLC_PEA	P13918 pisum sativ
11	738	20.8	439	1	GLCB_SOYBN	P25974 glycine max
12	699.5	19.7	410	1	VCLB_PEA	P02854 pisum sativ
13	697.5	19.7	463	1	VCL_VICFA	P08438 vicia faba
14	674	19.0	445	1	CANA_CANEN	P50477 canavalia e
15	667	18.8	445	1	CANA_CANGL	P10562 canavalia g
16	552	15.6	436	1	PHSB_PHAVU	P07219 phaseolus v
17	542	15.3	421	1	PHSB_PHAVU	P02853 phaseolus v
18	502.5	14.2	386	1	CVCB_PEA	P13919 pisum sativ
19	397	11.2	275	1	VCLA_PEA	P02855 pisum sativ
20	216.5	6.1	1898	1	CHAIN	Q07283 homo sapien
21	212	6.0	1407	1	TRHY_HUMAN	P37709 oryctolagus
22	194.5	5.5	124	1	TRHY_RABIT	P02856 pisum sativ
23	193.5	5.5	1549	1	VCL1_PEA	P22793 ovis aries
24	185	5.2	544	1	TRH_SHEEP	P24708 aotus trivi
25	179	5.1	499	1	INVO_AOTTR	P07730 oryza sativ
26	178.5	5.0	500	1	GLU2_ORYSA	P14614 oryza sativ
27	176.5	5.0	338	1	GLU5_ORYSA	P14594 oryza sativ
28	176.5	5.0	1023	1	LEGB_PEA	P33438 drosophila
29	176	5.0	585	1	GLT_DROME	Q08606 drosophila
30	175.5	5.0	1391	1	INVO_HUMAN	Q07676 homo sapien
31	174	4.9	495	1	MST2_DROHY	P07729 oryza sativ
32	174	4.9	499	1	GU12_ORYSA	P07728 oryza sativ
33	172	4.9	499	1	GU11_ORYSA	Q02898 oryza sativ
					GLUB_ORYSA	

ALIGNMENTS

34	172	4.9	887	1	YLX8_CAEEL	P46504 caenorhabdi
35	171.5	4.8	560	1	INVO_PANPA	P14591 pan paniscu
36	171	4.8	518	1	SSG2_AVESA	P14812 avena sativ
37	171	4.8	708	1	GBF_DICDI	P36417 dictyosteli
38	169	4.8	518	1	SSG1_AVESA	P12615 avena sativ
39	168.5	4.8	522	1	INVO_HYLLA	P17941 hylobates l
40	166.5	4.7	471	1	RU17_XENLA	P09406 xenopus lae
41	166.5	4.7	1240	1	YNJ1_YEAST	P53935 saccharomyc
42	165.5	4.7	285	1	INVO_CANFA	P18174 canis famil
43	165.5	4.7	1344	1	IF3A_MOUSE	P23116 mus musculu
44	165	4.7	255	1	LP61_EIMTE	P15714 eimeria ten
45	165	4.7	499	1	GLU4_ORYSA	P14323 oryza sativ

RESULT 1
VCLA_GOSHI

ID	VCLA_GOSHI	STANDARD;	PRT;	605 AA.
AC	P0799;
DT	01-MAR-1989 (Rel. 10, Created)
DT	01-MAR-1989 (Rel. 10, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)

DE   VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chian C.A., Borroto K., Kamalay J.A., Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XIX. Sequences and genomic organization of the alpha
RT   globulin (vicilin) genes of cottonseed.";
RL   Plant Mol. Biol. 9:533-546(1987).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC   BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC   CONVICILIN, CONGLYCININ, ETC.).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M19378; AAA33069.1; -.
DR   PIR; S06398; S06398.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL   1   23
FT   CHAIN   24   605   VICILIN GC72-A.
SQ   SEQUENCE   605 AA; 71049 MW; C9DB9371C97695 3B CRC64;

Query Match   30.8%;   Score 1092; DB 1; Length 605;
Best Local Similarity   36.4%;   Pred. No. 4e-60;
  Matches 250; Conservative 117; Mismatches 210; Indels 110; Gaps 17;

Qy   1 MAINTSNLCSLLFLLSLFLLSTTVSLAE------SEFD-RQEYEECKRQCMQLFTSGQMR 53
           ||| | || | | ||| |||| ||||| ||||||| ||||| | ||
Db   1 MVRNKSVFVLLF--SLFLSFGLLCSAKDFPGRRSEDDPQQRYEDCRKRC-QLETRGQTE 57

Qy   54 RCVSQOCDKRFEEDIDWSKYDNQEDPQTECQQCCRRCRQQESGPRQQQYCQRRCKEICEEE 113
           | | || | ||| | | | |::| | ||| |:|| | :|:| | ||| |::: : ::
Db   58 Q--DKCEDRSETQLKEEQORDGEDPQRRYQDCRQHCQQEERRLRP-----HCEQSCREQ 109

```
Qy  114  EEYNRQRDPQQQYEQCQQKHCQQRRETEPRHMQTCQQRCERRYEKEKRKQQKRYEEQQOREDE  173
         |:|:| ||||||:|:|  ||| :|| ||: |||::|
Db  110  YEKQQQQOPDKQFKECQQRCQWQEQORPERKQQCVKECREQYOEDPWKGER-------E  160

Qy  174  EKYEEERMKEEDNKRDPQQOREYEDCRRRCEQQOQEPRQQHQCQLRCREQQOROHGGRGGDMMNPQ  233
         |:| :||
Db  161  NKWREEEEES-----------------------------  171

Qy  234  RGGSGRYEGEEEEQSDNPYYFDERSLSTRFTEEGHISVLENFYGRSKLLRALKNYRLVL  293
         ||||| ||||:| :|
Db  172  -------DEGEQOQRNNPYYFHRRSFQOERFRREEHGNFRVLQRFADKIHILLRGINEFRIAI  224

Qy  294  LEANPNAFVLPTHLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLIN  353
         |||  |: |::  |  || |  ||||||
Db  225  LEANPNFVLPHHCDAEKIYVVTNGRGTVTVTHENKESYNVPGVVVRIPAGSTVYLAN  284

Qy  354  RDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF  413
         |
Db  285  QDNREKLTIAVLHRPVNNPGQFQKFFPAGQENPQSYLRIFSREILEAVFNTRSEQLDELP  344

Qy  414  G---------QOREGVIRASQEQIRELTRDDSESRIHWIIRGGESSRG-PYNLFNKRPLYS  465
             |:|
Db  345  GGRQSHRRQOGQGMFRKASQEQIRALSQGATSPR-----------GKGSEYAFNLLSQQTPRYS  397

Qy  466  NKYGQAYEVKPEDY-RQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMA  524
Db  398  NONGRFYEACRPNFQOQLREVDSSVVAFEINKGSIEVPHYNSKATFVVLVTEGNGHVEMV  457

Qy  525  CPHLSGRHGGRGGGKRHEEEEDV---------HYEQVRARLSKREAIVVLAGHPVFVSSGGN  577
         |:||:
Db  458  CPHLS-RQSSDWSSREEEEQEQEVERRSGQYKRVRAQLSTGNLFVVPAGHPVTFVASQN  516

Qy  578  ENLLFAFGI-NAQNNHENFLAGRERNVLQOIEPQAMELAFAAAPRKEVEESSFNSQOQSIF  636
         |:||
Db  517  EDLGLLGFGLLYNGQDNKRIFVAGKTNNV-ROWDRQAKELAFGVESSRLVDEVPNNNPQEESY  575

Qy  637  FPGPROHQOOSPRSTKQOOPLVSILDF  663
         ||| |||
Db  576  FVSGRDRRGFDER-RGSNNPLSPFLDF  601
```

RESULT 2
VCLB_GOSHI

```
ID   VCLB_GOSHI    STANDARD;    PRT;    588 AA.
AC   P09801;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS   Gossypium hirsutum (Upland cotton).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Malvales; Malvaceae; Gossypium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chlan C.A., Pyle J.B., Legocki A.B.; Dure L. III;
RT   "Developmental biochemistry of cottonseed embryogenesis and
RT   germination. XVIII. cDNA and amino acid sequences of the members of
RT   the storage protein families.";
RL   Plant Mol. Biol. 7:475-489(1986).
-!-  FUNCTION: SEED STORAGE PROTEIN.
-!-  SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
     BODIES.
-!-  SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
     CONVICILIN, CONGLYCININ, ETC.).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; M16891; AAA33071.1; -.
DR   PIR; A30838; FWCNAB.
DR   HSSP; P50477; 1CAX.
DR   INTERPRO; IPR00113; -.
DR   PFAM; PF00190; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL       1     25
FT   CHAIN       26    588       VICILIN C72.
SQ   SEQUENCE   588 AA;   69729 MW;  63E699B29ABADEB CRC64;

Query Match    30.6%;  Score 1082.5;  DB 1;  Length 588;
Best Local Similarity  38.5%;  Pred. No. 1.5e-59;
Matches  231;  Conservative 112;  Mismatches 182;  Indels 75;  Gaps 13;
```

```
Qy   76  EDPQTECQOQCQRCRQQESGPRQQOYCQRRCKEICEEEEEYNROR--DPQQOYEQCQKIC  133
         :||     |||:|| |     |       |||::|
Db   35  DDDPKRYEDCRRRCEWDTRGQKEQQOCEESCKSQYGEKDQOQQORHRPEDPQRRYEECQQEC  94

Qy  134  QRRETEPRHMQTCQQRCERRYEKRKQQKRYEEQQOREDEEKYEERMKEEDNKRDPQQORE  193
Db   95  --RQQEERQRPQCCQQRCLKRFEQEQQQ----------------------------SQRQ  123

Qy  194  YEDCRRRCEQQE--PRQOHQCOLRCREQQOROHGGRGGDMMNPQR------GGSGRYEGGE  245
         |::|            |||::|   ||
Db  124  FQECQOHCQHQQEQORPERKQQCVRECRERYQE----------NPWRREREEEAEEEETEEGBQ  175

Qy  246  EQSDNPYYFDERSLSTRFTEEGHISVLENFYGRSKLLRALKNYRLVLLEANNAFVLPT  305
         |||:||      :|  :|||||:||| ||
Db  176  EQSHNPFHFHRRSFQSRREEHGNFRVLQRFASRHPILRGINEFRLSILEANPNTFVLPH  235

Qy  306  HLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLINRDNNERLHIAKF  365
         | ||  | |  ||||   | :||||||
Db  236  HCDAEKIYLVTNGRGTLFLTHENKESYNVPGVVVRVPAGSTVYLANQDNKEKLIIAVL  295

Qy  366  LQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVFG---------QORE  418
             |||| ||:||||||   | :|| :|||||| ||
Db  296  HRPVNNPRQFEEFFPAGSQRPQSYLRAFSREILEPAFNTRSEQLDELFGGRQSHRRQOGQ  355

Qy  419  GVIIRASQEQIRELTRDDSESRIHWIIRGGESSRGPYNLFNKRPLYSNKYGQAYEVKPED  478
         |  |:|||||| |         :|    ||||  |
Db  356  GMFRKASQEQIRALSQEATSPR------EKSGE--RFAFNLLYRTPRYSNONGRFYEACPRE  409

Qy  479  YRQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMACHLSGRHGGRGGG  538
          | ||::|   || | :|||:||||||||::|  |  || |::|
Db  410  FROLSDINTVSALOLNQGSIFVPHYNSKATFVVLVNEGNGYVEMVSPHLPRQSSFEEEE  469

Qy  539  KRHEEEEDV-------HYEQVRARLSKREAIVVLAGHPVFVFVSSGCNENILLFAFG----I  587
         :|||::|            |::|   | | :|| |||||:| |    |
Db  470  EQOQEQOEOEERRSGQYRKIRSQLSRGDIFVVPANFPVTFVASQNONLRMTGFGLYNQNI  529

Qy  588  NAQNNHENFLAGRERNVLQOIEPQAMELAFAAAPRKEVEESSFNSQODQSIFFFPGRPRQHQQOS  647
         |  |:|  ||||::|  |  ||||||||   | | |  || ||
Db  530  NPDHNQRIFVAGKINHV-ROWDSQAKELAFGVSSRLVDEIFNNNPQESYFVS-RQRQRAS  587
```

RESULT 3
GLB1_MAIZE

```
ID   GLB1_MAIZE    STANDARD;    PRT;    573 AA.
AC   P15590;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).
GN   GLB1.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. INBRED LINE VA26;
RA   Belanger F.C., Kriz A.L.;
```

```
RT   "Molecular characterization of the major maize embryo globulin encoded
RT   by the Glb1 gene.";
RL   Plant Physiol. 91:636-643(1989).
RN   [2]
RP   SEQUENCE OF 87-100.
RX   MEDLINE=89374022; PubMed=2775172;
RA   Kriz A.L.;
RT   "Characterization of embryo globulins encoded by the maize Glb
RT   genes.";
RL   Biochem. Genet. 27:239-251(1989).
-!-  PTM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC       MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
-!-  POLYMORPHISM: THE THREE MOST COMMONLY OCCURING GLB1 ALLELES HAVE
CC       THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC       PROTEINS, RESPECTIVELY.
-!-  SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```
DR   EMBL; M24845; AAA33467.1; -.
DR   HSSP; P50477; 1CAW.
DR   MAIZEDB; 30181; -.
DR   INTERPRO; IPR001113; -.
PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
SIGNAL       1     18
PROPEP      19     86          OR 21 (POTENTIAL).
CHAIN       87    573          GLOBULIN-1 S ALLELE.
CARBOHYD   349    349          N-LINKED (GLCNAC) (POTENTIAL).
SEQUENCE   573 AA;  65029 MW;  525ED1D00A062976 CRC64;
```

```
Query Match          25.6%;
Best Local Similarity 39.9%;
Matches  201;  Conservative  91;  Mismatches 180;  Indels  32;  Gaps  10;
Score 905;  DB 1;  Length 573;  Pred. No. 1.1e-48;

Qy  182  EEDNKRDPQQQREYEDCRRRCEQQOEPRQQHQCQLRCRCEDQQROHGRGGDMMN-PQRGCSG--  238
              |:||                  |  |||   |||    |:   |  |:   |  :  |:   |:  |
Db   25  EDDNHHHHGGHHKSGRCVRRCEDRPWHQRPRCLEQCREEERKRQERSRHEADDRSGEGSS    84

Qy  239  -----RYEEGEEEQSD-NPYYEFDERSLSTRFFTEEGHISVLENFYGRSKLLRALKNYRLV   292
              :  |||   ::||    : |   |  || :|| |: |:|: |:|| | | |:| ||
Db   85  EDEREREQEKEEKEQKDRRPYVEFDRRSFRRVVRSEQGSLRVLRPFFDEVSRLLRGIRDYRVA  144

Qy  293  LLEANPNAFVLPTHLDADAILLVIGGRGALKMIHHDNRESTNLECGDVIRIPAGTTFYLI   352
              :|||||||  :|:|:||  | |:   ||||||| |:: | |:|  |  | |||| | ||
Db  145  VLEANPRSFVVPSHTDAHCIGYVAAEGEGVVTTIENGERRSYTIKQGHVFVAPAGAVTYLA  204

Qy  353  NRDNENRLHIAKFLQTTISTPGQYKEFFPAGGQNPEPYLSTESKEILEAALNTQTEKLRGV  412
              | |  | :: || | |: ::  : ::    |:|  |  |:  |  |  |:| || ||
Db  205  NTDGRKKLVITKILHTISVPGEFQFFFGPGGRNPESFLSSFSKSIQRAAYKTSSDRLERL   264

Qy  413  FGQ--QREGVIIRASQEQIRELTRDDSE---SRHWHIRRGGESSSRGPYNLFNKRPLYSNK  467
              |: |  : | |  | :|:| |:| |     :  :: | | | ||  | | |: |: |:
Db  265  FGRHGQQDKGIIVRATEQTRELRRHASEGGHGPHWPLPPFGE-SRGPYSLLDQRPSIANQ   323

Qy  468  YGQAYEVKPEDYRQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMACPH   527
              |: |  :| ::: : ||  || : || |  | |:|  ||:||||||:|:  :|| |  |
Db  324  HGQLYEADARSPHDLAEHDVSVSFANITAGSMSAPLYNTRSFKIAYYPNGKGYAEIVCPH   383

Qy  528  LSGRHG----GRGGGKRHEEEDVHYEQ------                              574
              | ||||     :| |  || : |
Db  384  RQSQGGESERERGKGRRSSEEEEEESSEEQEEVGQGYHTIRARLSPGTAFVVPAGHPEVAVA 443

Qy  575  SGNENLLLFAFGINAQNNHENFLAGRERNVLQQIEFPQAMELAFAAPRKEVEESFNSQDQS  634
              | ||:|| | | || | ||  ||  |:  :: |   |    | |  : |   :  |
Db  444  SRDSNLQIVGFPEVHADRNEKVFLAGAD-NVLQKLDRVAKALSFASKAEEVDEVLGSRREK  502
```

RESULT 4

GLCX_SOYBN

```
ID   GLCX_SOYBN     STANDARD;      PRT;   639 AA.
AC   P11827;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   BETA-CONGLYCININ, ALPHA' CHAIN PRECURSOR.
GN   CG-1.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83143289; PubMed=6897678;
RA   Schuler M.A., Ladin B.F., Pollaco J.C., Freyer G., Beachy R.N.;
RT   "Structural sequences are conserved in the genes coding for the
RA   alpha, alpha' and beta-subunits of the soybean 7s seed storage
RT   protein.";
RL   Nucleic Acids Res. 10:8245-8261(1982).
RN   [2]
RT   "The glycosylated seed storage proteins of Glycine max and Phaseolus
RA   vulgaris. Structural homologies of genes and proteins.";
RT
RL   J. Biol. Chem. 261:9228-9238(1986).
-!-  FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC       DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC       CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
-!-  SUBUNIT: THE ALPHA', ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC       VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
-!-  SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
-!-  SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```
DR   EMBL; M13759; AAB01374.1; -.
DR   EMBL; J01290; -; NOT_ANNOTATED_CDS.
DR   PIR; B24810; B24810.
DR   HSSP; P02853; 2PHL.
DR   INTERPRO; IPR001113; -.
PFAM; PF00546; Seedstore_7s; 2.
KW   Seed storage protein; Signal; Glycoprotein; Multigene family.
SIGNAL       1     22
PROPEP      23     62
CHAIN       63    639          BETA-CONGLYCININ, ALPHA' CHAIN.
CARBOHYD   277    277          N-LINKED (GLCNAC) (POTENTIAL).
CARBOHYD   551    551          N-LINKED (GLCNAC) (POTENTIAL).
CONFLICT   543    543          P -> L (IN REF. 2).
CONFLICT   549    549          M -> V (IN REF. 2).
CONFLICT   608    608          S -> T (IN REF. 2).
SEQUENCE   639 AA;  74325 MW;  469BF24C79651E3F CRC64;
```

```
Qy  635  IFFPGPRQ---HQQQSPRSTKQQQ   655
              |:||       ||||
Db  503  GFLPGPKESGGHEEEREQEEEEREE  526
```

```
Query Match   24.2%;   Score 856.5;   DB 1;   Length 639;
```

RESULT 5

```
Best Local Similarity 30.4%;  Pred. No. 1.2e-45;
Matches 210; Conservative 154; Mismatches 221; Indels 109; Gaps 19;

ID   AH12_ARAHY     STANDARD;      PRT;   626 AA.
AC   P43238;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ALLERGEN ARA H 1, CLONE P41B (ARA H I).
OS   Arachis hypogaea (Peanut).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Arachis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. FLORUNNER;
RX   MEDLINE=96013631; PubMed=7560062;
RA   Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT   "Recombinant peanut allergen Ara h I expression and IgE binding in
RT   patients with peanut hypersensitivity.";
RL   J. Clin. Invest. 96:1715-1721(1995).

Qy    11 LLFLLSLFLLSTTVSLAEESEFDRQ--EYEEECKRQCMQLETSCGQMRRCVSQCDKRFEEDID 68
Db     8 LLLGVVFLASVSVSFGIAYWEKQNPSHNKCLRSCNSEKDSYRNQCHARCNLLKVEE-- 65

Qy    69 WSKYDNQEDPQTECQQCQ-RRCRQQESGPRQQQYCQQRCKEICEEEE------EYNRQR 120
Db    66 -EEECEEGQIPRRPQHPEREPRQQHGEK-- 109

Qy   121 DPQQ--QYEQCQKHCQRRETEPRHMQTCQQRCCERRYEKEKRKQQKRYEEQQREDEEKYEE 178
Db   110 QPHQEEHEQKEEHEWHRKEEKHGGKGSEE-----EQDEREHPRPHQPHQKE-EEKHEW 162

Qy   179 RMKEEDNKRDPQQREYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRGCSG 238
Db   163 QHKQEKHQGKKESEEEEED-----QDEDEDQ 198

Qy   239 RYEEGEEEQSDNPYIFDERSLSTRFRTEEGHISVLENEYGRSKLLRALKNYRLVLLEANP 298
Db   199 SQREPRRHKNKNPFHFNSKRFQTLFKNQYGHVRVLQRFNKRSQQLQNLRDYRILEFNSKP 258

Qy   299 NAFVLPFHLDADILLVIGGRGALKMIHHHDNRESYNLECGDVIRIP------ 344
Db   259 NFLLLPHHADADYLIVILLNGTALTLLVNNDDRDSYNLQSGDALRVPAGTTFYVVNPDNDE 318

Qy   345 ------AGTTFYLINRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILE 399
Db   319 NLRMIAGTTFYVVNPNDENLRMTTLAIPVNKPGRFESEFFLSSTQAQSYLQGFSKNILE 378

Qy   400 AALNTQTEKLRGVF------GQQ------REGVIIRASQEQIRELTRDDSESRHWHIRRGG 448
Db   379 ASYDTKFEEINKVLFGREEQQOGEERLQESVIVEISKKQIRELSKHAKSSS---RKTI 434

Qy   449 ESSRGPYNLFNKRPLYSNKYGQAYEVKPEDYRQLQDMDLSVFIANVTQGSMMGPFFNTRS 508
Db   435 SSEDKPFNLGSRDPIYSNKLGLKLFEITQRN-PQLRDLDVFLSVDMNEGALFLPHFNSKA 493

Qy   509 TKVVVVASGEADVEMACPHLSGRHGGRGGGGKRHEEEED--VHYEQVRARLSKREAIVVLAG 567
Db   494 IVVLLVINEGEANIELV---------GIKEQQORQQOEEQPLEVRKYRAELSEQDIFVIPAG 545

Qy   568 HPVVFVSSGNENLLFAFGINAQNHHENFLAGRERNVLQQIEPQAMELAFAAPRKEVEES 627
Db   546 YPVMV--NATSDLNFAFGINAENNQRNFLAGSKDNVISQIPSQVQELAFPRSAKDIENL 603

Qy   628 FNSQDQSIFFPGPRQHQQOQSPRSTKQQQPLVSIL 661
Db   604 IKSQSESYFVDA--QPQQKEEGNKGRKGPLSSIL 635
```

```
CC   -!- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
DR   EMBL; L34402; AAB00861.1; -.
DR   HSSP; P50477; 1CAW.
DR   INTERPRO; IPR001111; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Allergen.
SQ   SEQUENCE   626 AA;  71345 MW;  1A6BBBE41490D0E3 CRC64;

Query Match 24.0%;  Score 849.5;  DB 1;  Length 626;
Best Local Similarity 34.4%;  Pred. No. 3.2e-45;
Matches 212; Conservative 108; Mismatches 204; Indels 93; Gaps 20;

Qy    93 ESGPRQQQY--CQRRCKEICEEEEEYNQRDPQQQYEQCQKHCQRRETEPR------H 142
Db    26 KSSPYQKKTENPCAQRCLQSCQQOEPDDLKQK-----ACESRCTKLEYDPRCVYDPRGH 78

Qy   143 MQTCQQRC----ERR------YEKEKRKQQKRYEEQQREDEEKYEERMKEEDNKRDPQQR 192
Db    79 TGTTNQRSPPGERTRGRPQDYDDRR--QPRREDGGRWQPAGPRERERREED-WRQPR-- 133

Qy   193 EYEDCRRRCEQQEPRQQHQCQLRCREQQRQHGRGGDMMNPQRGCSGRYEEGEEEQSDNPY 252
Db   134 --EDWRRPSHQQ-PR-----KIRPEGREGE----QEWGTPGSHVREETSRNNPF 175

Qy   253 YFDERSLSTRFRTEEGHISVLENEYGRSKLLRALKNYRLVLLEANPNAFVLPFHLDADAI 312
Db   176 YFPSRRFSTRYGNQNGRIRVLQRFDQRSRFQFONLQNHRIVQIEAKPNTLVLPKHADADNI 235

Qy   313 LLVIGGRGALKMIHHHDNRESYNLECGDVIRIPAGTTFYLINRDNNERLHIAKFLQTISTP 372
Db   236 LVIQQGQATVTVANGCNNRKSFNLDEGHALRIPSGFISYLNRHDNQNLRVAKISMPVVNTP 295

Qy   373 GQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF----------GQOR--- 417
Db   296 GQFEDFFPASSRDQSSYLQGFSRNTLEAAFNAEFNEIRRVLLEENAGGEOEERGQRRST 355

Qy   418 -----EGVIIRASQEQIRELTRDDSESRHWHIRRGGESSRG---PYNLFNKRPLYSNK 467
Db   356 RSSENNEGVIVKVSKEHVEELTKHAKS----VSKKGSEEEGDITNPINLREGGEPDLSNN 410

Qy   468 YGQAYEVKPEDYR-QLQDMDLSVFIANVTQGSMMGPFFNTRSKVVVVASGEADVEMACP 526
Db   411 FGKLFEVKPDKKNPCLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVNKGTGNLELVAV 470

Qy   527 HLSGRHGRGGGKRHEEEEDVHYE-----QVR----ARLSKREAIVVLAGHPVVFVSSGNEN 579
Db   471 RKEQQORCGR---REEEEDEDEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSELH 527

Qy   580 LLLFAFGINAQNHHENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESPNSQDQSIFFPG 639
Db   528 LL--GFGINAENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKNQKESHFVSA 585

Qy   640 PRQHQQOQSPRSTKQQQP 656
Db   586 RPQSQSQSPSSPSPEKESP 602
```

RESULT 6

```
ID   AH11_ARAHY     STANDARD;      PRT;   614 AA.
AC   P43237;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
```

```
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ALLERGEN ARA H 1, CLONE P17 (ARA H I).
OS   Arachis hypogaea (Peanut).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Arachis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. FLORUNNER;
RX   MEDLINE=96013631; PubMed=7560062;
RA   Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT   "Recombinant peanut allergen Ara h I expression and IgE binding in
RT   patients with peanut hypersensitivity.";
RL   J. Clin. Invest. 96:1715-1721(1995).
-!- SIMILARITY: TO 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
-!- CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; L38853; AAA60336.1; -.
DR   HSSP; P50477; 1CAW.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Allergen.
SQ   SEQUENCE   614 AA;  70283 MW;  1DDACF217EEC5F31 CRC64;

Query Match             34.7%;  Score 849;  DB 1;  Length 614;
Best Local Similarity   34.7%;  Pred. No. 3.3e-45;
Matches  219;  Conservative  110;  Mismatches  216;  Indels  86;  Gaps  21;

Qy    70 SKYDNQEDPQTEC-QQCQORCRQQESGPRQQQYCQRRCKEICEEEEYNRQRDPQQQYEQ 128
            :|:| .| :|        | |::|    :|| : ::     :      |   ||
Db    27 SPYRKTENP---CAQRCLQSC-QQEPDLIKQKACESRCTKL-----EY----DPRCVYD- 72

Db    73 ------TGATNQRHPPGERTRGRPGDYDDRRQPRREGGRWGPAEPEREEREED-WRQ 125
              |  | .: |: ||    ||           ||        |||||  ||| |
Qy   129 CQXHCQRRETEPRHMQTCQQCQRCERRYEKEKRKQQKRYEEEQQREDEEKYEERMKEEDNKRD 188

Qy   189 PQQREYEDCRRRCQEQEPRQOHQCQLRCREQQRQHGRGGDMMNPQRGGSGRYEEGEEEQS 248
            | || :| :|          |     :| | :         ::|  :::    |
Db   126 PR-----EDWRRPSHQQ-PR----------KIRPEGREGE----QEWGTPGSEVREETSR 165

Qy   249 DNPYFDERSLSTRFRTEEGHISVLENPYGRSKLLRALKNYRLVLLEANPNAEVLPTLHLD 308
           :||:|  | ||:|: | :    |:|  |:|||    :::|: |: || ||: ||| :| |
Db   166 NNPFYPSRREFSTRYGNGRIRVLQREFDQRSKQFQNLQNHRIVQIEARPNTLVLPKHAD 225

Qy   309 ADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLINRDNNERLHIAKFLQT 368
           || :|||  | |  : : :|:| |: |: |  ||||:|: ||:|:|  ::|  ||||: |
Db   226 ADNILVIQQGQAIVTVANGNNRKSFNILDEGHALRIPSGFISYILNRHDHQNLRVAKISMP 285

Qy   369 ISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF------GQQ 416
           :.|||: |:|||| . :  ||| |||:||:||| |: | :|  |        |:
Db   286 VNTPGQFEDFFPASSRDQSSYLQGFSRNTLEAAFNAEFNEIRVLLEENAGGEQEEQR 345

Qy   417 R---------EGVIIRASQEQIRELTRDDSESRHWIHIRGGESS--RGPYNLENKRPLYSN 466
           |         |||::|.| || ||||   :::           :  ||| |  : ||||
Db   346 RRSTRSSDNEGVIVKVSKEHVQELTKHAKSVS---KKGSEEDITNPINLRDGEPDLSN 401

Qy   467 KYGQAYEVKPEDYR-QLQDMDLSVFIANVTQGSMGPFNTRSTKVVVASGEADVEMAC 525
            :|: :|||||: : |||| || : ::||. |: | |:|:| :|:|:. |  ::||
Db   402 NFGRLFEVKPDKKNPQLQDLDMMLTCVEIKEGALMLPHFNSKAMVIVVVNKGTGNLELVA 461

Qy   526 PHLSGRHGGRGGKRHEEEEDVHYE-----QVR---ARLSKREAIVVLAGHPVFVFSSGNE 578
            | || :| :|| ||:||||  |       |:    |||  |  | :||||||:  |:|:|
Db   462 VRKEQQORGRREQEWEEEEGSNREVRRYTARLKEGDVFIMPAAHPVAINASSEL 521

Qy   579 NLLEAFGINAQNHIENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESFNSQDQSIFFP 638
            :||.|||||::.|| || |  ::|::|||| |.: ||  .| ||  :|: ||:  :
Db   522 HLL--GFGINAENNHRIFLAGDKDNVIDQIEKQAKDIAFPGSGEQVEKLIKNQRESHFVS 579

Qy   639 G-PRQHQQOQSPRSTKQQQ------PLVSIL 661
           | |: | |  |  : |:||       |:|||
Db   580 ARPQSQSPSSPEKEDQEEENGGGKGPLLSIL 610


RESULT 7
GLCA_SOYBN

ID   GLCA_SOYBN     STANDARD;      PRT;   605 AA.
AC   P13916;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COTYLEDON;
RX   MEDLINE=91355860; PubMed=2103438;
RA   Sebastiani F.L., Farrel L.B., Schuler M.A., Beachy R.N.;
RT   "Complete sequence of a cDNA of alpha subunit of soybean beta-
RT   conglycinin.";
RL   Plant Mol. Biol. 15:197-201(1990).
-!- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
-!- DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
-!- CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
-!- SUBUNIT: THE ALPHA-', ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
-!- VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
-!- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
-!- VACUOLAR PROTEIN BODIES.
-!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
-!- CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; X17698; CAA35691.1; -.
DR   PIR; S14681; FWSYBA.
DR   HSSP; P50477; 1CAW.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal; Glycoprotein; Multigene family.
FT   SIGNAL        1     22
FT   PROPEP       23     62
FT   CHAIN        63    605       BETA-CONGLYCININ, ALPHA CHAIN.
FT   CARBOHYD    261    261       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   605 AA;  70293 MW;  CBEBA30506BBBC57 CRC64;

Query Match             30.5%;  Score 840.5;  DB 1;  Length 605;
Best Local Similarity   30.5%;  Pred. No. 1.1e-44;
Matches  205;  Conservative  144;  Mismatches  223;  Indels  101;  Gaps  17;

Qy    11 LLFLFLSLFLLSTTVS--LAESEFDRQEYEECKRQCMQLETSGQMRRCVSQCDKRFEEDID 68
           ||: .|.|| |::.|:  |   :.:|:||  ::|   | ::|. ::::| |.:|  ::  |
Db     8 LLILGLVFLASVSVSFGIAYWEKENPKIHNKCLQSCNSERDSYRNQACHARC-----N 59

Qy    69 WSKYDNQEDPQTECQQCQORCRQQESGPRQQQYCQRRCKEICEEEEYNRQRD---PQQQ 125
           :||:|:|:|     | ||  :|| :  |  |||      ||| ||:| :|:|   |
Db    60 LLKVEKECCEGEIPRPRPRPQHPEREPQQPGE-------KEEDEDQPRPIPFPRPQ 110
```

```
Qy  126 YEQCQKHCQRRETE-PRHMQTCQQRCERRYEKEKRKQQKRYEQQQRDEEKYEERMKEED 184
            :|||: :| :||: ||: |||:|| | :: ||| ||||
Db  111 PRQEEHEQREEQEWPR--------KEEKRGEKGSEEEDEDEDEEQDER--QFP 154

Qy  185 NKRDPQQREYEDCRRRCEQQEPRQQHQCCQLRCREQQOHGRGGDMMNPQRGGSGRYEGE 244
Db  155 FPRPPHQKE-------ERNEEDEDE------BQQRESEESED------SELR 188

Qy  245 EEQSDNPYYFDERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLP 304
Db  189 RHKNKNPELFGSNRFEELFKNQYGRIRVLQRFNQRSPQLQNLRDYRILEFNSKPNTLLLP 248

Qy  305 THLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLLINRDNNERLHIAK 364
Db  249 NHADADYLIVILNGTAILSLVNNDRDSYRLQSGDALRVPSGTTYYVVVNPDNNENLRLIT 308

Qy  365 FLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF---------GO 415
Db  309 LAIPVNKPGRFESFFLSSTEAQQSYLQGFSRNILEASYDTKFEEINKVLFSREEGQQQGE 368

Qy  416 QR--EGVIIRASQEQIRELTRDDSESRHWHIRRGGESSSRGPYNLFNKRPLIYSNKYGQAYE 473
Db  369 QRLQESVIVEISKEQIRALSKRAKSSS----RKTISSEDKPENLRSRDPIYSNKLGKFFE 424

Qy  474 VKPEDYRQLQDMDLSVFIANVTQGSMMGPFNTRSTKVVVASGEADVEMACPHLSGRHG 533
Db  425 ITPEKNPQLRDLDIFLSIVDMNEGALLLPHFNSKAIVILVINEGDANIELV------- 475

Qy  534 GRGGGK------RHEEEEDVHYEQVRARLSKREAIVVLLAGHPVVFVSSGNENLLLFAFGIN 588
Db  476 --GLKEQQOEEQQEQPLEVKRYRAELSEQDIFVIPAGYPVVV--NATSNLNFFAIGIN 530

Qy  589 AQNNHENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESFNSQDQSIFFPGPRQHQQQSP 648
Db  531 AENNQRNFLAGSQDNVISQIPSQVQELAFPGSAQAVEKLLKNQRESYFVDAQPKKKEEGN 590

Qy  649 RSTKQQQPLVSIL 661
Db  591 KGRK--GPLSSIL 601


RESULT 8

ID   CVCA_PEA     STANDARD;     PRT;     571 AA.
AC   P13915;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   CONVICILIN PRECURSOR.
GN   CVCA.
OS   Pisum sativum (garden pea).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Pisum.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=CV. FELTHAM FIRST;
RX   MEDLINE=88326208; PubMed=3415641;
RA   Bown D., Ellis T.H.N., Gatehouse J.A.;
RT   "The sequence of a gene encoding convicilin from pea (Pisum sativum
RT   L.) shows that convicilin differs from vicilin by an insertion near
RT   the N-terminus.";
RL   Biochem. J. 251:717-726(1988).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X06398; CAA29695.1; -.
DR   PIR; S00566; S00566.
DR   HSSP; P02853; 2PHL.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Multigene family; Signal.
FT   SIGNAL        1     28
FT   CHAIN        29    571       CONVICILIN.
SQ   SEQUENCE   571 AA;  66989 MW;  749CFBEB2D16D57B CRC64;

Query Match             23.2%;  Score 821.5;  DB 1;  Length 571;
Best Local Similarity   31.0%;  Pred. No. 1.5e-43;
Matches  211;  Conservative  112;  Mismatches  205;  Indels  153;  Gaps  19;

Qy   11 LLFLLSLFLLSTTVSLAESEFDRQEYEDCKRQCMQLETSGQMRCVSQCDKFEEID1DW- 69
         ||||   :||| |||: ||: : |  :|:
Db   12 LLFLGIIFLASVCVTYA-------NYDEGS----ETRVPGQRERGRQEGEKERHGEWR 60

Qy   70 SKIDNQEDFDQTECQQCQRRCRQQESGPRQQQYCQRRCKEICEEEEEYNRQRDPQQQYEQC 129
                                    ::|
Db   61 PSYEKEEHE-------------------------EEKQKYRYQREKKEQ---- 84

Qy  130 QKHCQRRTEPRHMQTCQQRCERRYEKEKRKQQKRYEQQQRDEEKYEERMKEEDNKRDP 189
           :: | || |                   || :  ||
Db   85 -----KEVQPG-------RERWERE-------EDEEQVEEWRGSQRREDP 116

Qy  190 QQREYEDCRRRCEQQEPRQQHQCCQLRCREQQOHGRGGDMMNPQRGGSGRYEGEEEQSD 249
         :|  :                           |  ||| |
Db  117 EER------ARLRHREERTKRDR-----RHQREGEEEERSSESQEHR 152

Qy  250 NPYYFDERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLPTHLDA 309
         ||          |: :|| | : | :   |  |:||:||: || | |  |   : |:|||
Db  153 NPFLPFKSNKFLTLFENENGHIRRLQRFDKRSDLPENLQNYRLVEYRAKPHTIFLPQHIDA 212

Qy  310 DAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLLINRDNNERLHIAKFLQTI 369
         | ||:|: | | : | ::|||:|||| ||:|:|||||| || : : : ||| : : | :
Db  213 DLILVVLNGKAILTVLSNDRNSYNLERGDTIKIPAGTTSYLVNODDEEDLRVVDFVIPV 272

Qy  370 STPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF------------ 413
           ||  |  |  | :| : |||||:||| ||||:|| :| | |
Db  273 NRPGKFEAF--GLSENKNQYLRGFSKNILEASLNTKYETIEKVLLEEQKIKPQQLRDRKR 330

Qy  414 --GQQREQVIIRASQEQIRELTRDDSESRHWHIRRGGESSSRGPYNLFNKRPLIYSNKYGQ 470
           | |: : :|| ||:|| ||| :|| |        |   | ||: |:||  |:||||:|
Db  331 TQQGEERD-AIIKVSREQIEELRKLAKSSS-----KKSLPSEFEPPNLRSHKPEYSNKFGK 385

Qy  471 AYEVKPE-DYRQLQDMDLSVFIANVTQGSMMGPFNTRSTKVVVASGEADVEMACPHLS 529
         |  ||  : :|||||:|:|| |:|: |:| :||  : || || | | ::   || |||
Db  386 LFEITPEKKYPQLQDLDILVSCVEINKGALMLPHYNSRAIVLJVNECKGNLELL---- 440

Qy  530 GRHGGRGGGKRHEEEE------------DVHYEQVRARLSKREAIVVLLAGHPVVFVSSGNENLLL 582
          : | |||||:|::            :   ::|||| || |||:| || ||||  :|||| |
Db  441 -----------GLKNEQQEREDRKERNNEVQRYEARLSPGDVVIIPAGHPVAISASSNLNL- 491

Qy  583 FAFGINAQNNHENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESFNSQDQSIFFPG-PR 641
         | ||||: ::::|:||| :|||:|:||:|||| | ||  :| | | :| |   ||| |
Db  492 -GFGINAKKNQRNFLSGSQDNVISQIENPVKELTFPGSSQQEVNRLIKNQKQSHFASAEPE 550

Qy  642 QHQQQSPRSTKQQQPLVSIlD 662
         |::::| |:  |  ||||: ||
Db  551 QKEEESQR---KRSPLSSVLD 568


RESULT 9

ID   SBP_SOYBN     STANDARD;     PRT;     524 AA.
AC   Q01672;
DT   01-JUN-1994 (Rel. 29, Created)
```

```
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   SUCROSE-BINDING PROTEIN PRECURSOR (SBP).
GN   SBP.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-50.
RC   TISSUE=EMBRYO;
RX   MEDLINE=93104680; PubMed=1467654;
RA   Grimes H.D., Overvoorde P.J., Ripp K., Franceschi V.R., Hitz W.D.;
RT   "A 62-kD sucrose binding protein is expressed and localized in
RT   tissues actively engaged in sucrose transport.";
RL   Plant Cell 4:1561-1574(1992).
CC   -!- FUNCTION: PLAYS A ROLE IN SUCROSE TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC   -!- TISSUE SPECIFICITY: ASSOCIATED WITH THE PLASMA MEMBRANE OF THE
CC       SEVERAL CELL TYPES ENGAGED IN SUCROSE TRANSPORT, INCLUDING THE
CC       MESOPHYLL CELLS OF YOUNG SINK LEAVES, THE COMPANION CELLS OF
CC       MATURE PHLOEM AND THE CELLS OF DEVELOPING COTYLEDONS.
CC   -!- DEVELOPMENTAL STAGE: IN THE COTYLEDON, EXPRESSION IS NOT DETECTED
CC       UNTIL 10 DAYS AFTER FERTILIZATION. BETWEEN 10-19 DAYS AFTER
CC       FERTILIZATION, EXPRESSION INCREASES RAPIDLY BUT DECLINES 20-30
CC       DAYS AFTER FERTILIZATION. 30 DAYS AFTER FERTILIZATION, NO
CC       EXPRESSION OCCURS. THIS EXPRESSION PATTERN CLOSELY PARALLELS THE
CC       RATE OF SUCROSE UPTAKE IN THE COTYLEDON.
```

```
DR   EMBL; L06038; AAB03894.1; -.
DR   PIR; JQ1730; JQ1730.
DR   HSSP; P50477; 1CAW.
DR   InterPro; IPR001113; -.
DR   Pfam; PF00546; Seedstore_7s; 1.
KW   Transport; Sugar transport; Signal; Membrane.
FT   SIGNAL        1     29
FT   CHAIN        30    524       SUCROSE-BINDING PROTEIN.
SQ   SEQUENCE   524 AA;  60522 MW;  0251EE50796F341 CRC64;

Query Match             22.8%;  Score 808;  DB 1;  Length 524;
Best Local Similarity   33.9%;
Matches  194;  Conservative  97;  Mismatches 179;  Indels 102;  Gaps 17;

Qy  105  RCKEICEEEENYNRQRDEQ--QQYEQCOQKHCQORRETEPRHMQTCQQRCERRYEKEKRRQQ  162
         ||| ||       :| |:||  ||| :    |   |:|  |:   |:|  |::|::
Db   27  KCKETEVEE------------DPELVTCKHQCCQQQQYTEGDKR---VCLQSCD-RYHRMKQERE  76

Qy  163  KRYEEQQOREDEEKYEEEMKEEDNKRDPQQOREYEDCRRRCEOQOEPRQOHQCOLRCREQOQRQ  222
         |||                                                    EQOEQ
Db   77  KQIQEETREKKEE-ESREREE------------------------------------  101

Qy  223  HGRGGDMMNPQRGGSGRYEEGEGEEQSDNPYYFDE-RSLSTRFRTEEGHISVLENFYGRSK  281
         |                     :|||||||:|::   :||  |::|:||:||:||:||:
Db  102  H----------EEQQENPYIFEEDKDFETRVETEGGRIRVLKKFTEKSK  140

Qy  282  LLRALKNYRLVLLEANDNAFVLPTHLDADALLLVIGRGALKMIHIHDNRESYNLECGDVI  341
         |||:| || |:||:||:| ||:|:|:|:|:|| |||||||||::|:|:||||| |:|:|
Db  141  LLQGIENFRLAILEARAIHTFVSPRHFDSEVFNIKGRAVLGLVSESETEKITLEPGDMI  200

Qy  342  RIPAGTTFYLINRDNNERLHIA--KFLQTISTPGQYKEEFFAGGQNPEPYLSTFSKEILE  399
         :|||||.:|::|||:|::| :   |||:|||||:||||||||:||:|||:|| |:|:|:|
Db  201  HIPAGTPLYIVNRDENDKLFLAMLHIPVSVSTPGKFEEFFAPGGRDPESVLSAFSWNVLQ  260

Qy  400  AALNTQTEKLRGVFGQOREGVIIRASQEOIRELTRDDSESRHWHIRRGGESSRGPYNLFN  459
```

(continued — right column)

```
Db  261  AALQTPKGKLENVFDQQNEGSIFRISREQVRALA-PTKKSSWWPF--GGE-SKPQFNIFS  316
         |||  | ||||||| |:::|||:||||:||||||  |||:|| ||   || :|||:||:|
Qy  460  KRPLYSNKYGQAYEVKPEDYRQ--LQDMDLSVFIANVTQGSMMGPFNFRSTKVVVASG  517
         |||:||:|||: ||   |:||   | |:| |::| :||:|| | || :|| |::::|
Db  317  KRPISNGYGRLTEVGPDDDEKSWLQRLNLMLTFTNITQRSMSTIHYNSHATKIALVIDG  376

Qy  518  EADVEMACPILSGRHGGRGGGKRHEEEEDVHYEQVRARLSKREAIVVLLAGHPVVFVSSGN  577
         : :|::.||::|       |:|||:    :|| |||:|||:||:::::||||| |::|:||
Db  377  RGHLQISCPHMS------SRSSHSKHDKSSP-SYHRLSSDLKPGMVFVVPPGHPFVTIASNK  431

Qy  578  ENLLFAFGINAQNHHENFLAGRERNVLQOIEPQAMELAFAAPRKEVEESF----------  628
         ||||  | :||| ::  || ||:::|::: |  | :|||:|   |:|
Db  432  ENLLMICFEVNARDNKKFTFAGKD-NIVSSLDNVAKELAFNYPSEMVNGVFLLQRFLERK  490

Qy  629  ----------NSQDQSIFFPGPROHQQQSPRS  650
                   ||:| |:|||||| |::||
Db  491  LIGRLYHLPHKDRKESFFFPELPREERGRRA  522
```

```
RESULT 10
ID   VCLC_PEA               STANDARD;     PRT;   459 AA.
AC   P13918;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   VICILIN PRECURSOR.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Pisum.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEEDLING;
RA   Higgins T.J.V., Newbigin E.J., Spencer D., Llewellyn D.J., Craig S.;
RT   "The sequence of a pea vicilin gene and its expression in transgenic
RT   tobacco.";
RL   Plant Mol. Biol. 11:683-695(1988).
RN   [2]
RP   SEQUENCE OF 27-459 FROM N.A. (CLONE PDUB9).
RC   STRAIN=CV. FELTHAM FIRST;
RX   MEDLINE=88326226; PubMed=3046604;
RA   Watson M.D., Lambert N., Delauney A., Yarwood J.N., Croy R.R.D.,
RA   Gatehouse J.A., Wright D.J., Boulter D.;
RT   "Isolation and expression of a pea vicilin cDNA in the yeast
RT   Saccharomyces cerevisiae.";
RL   Biochem. J. 251:857-864(1988).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- MISCELLANEOUS: THERE ARE AT LEAST 11 GENES FOR VICILIN SUBUNITS.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; X14076; CAA32239.1; -.
DR   EMBL; Y00722; CAA68708.1; -.
DR   PIR; S00567; S00567.
DR   PIR; S08505; S08505.
DR   HSSP; P50477; 1CAW.
DR   InterPro; IPR001113; -.
DR   Pfam; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Multigene family; Signal.
FT   SIGNAL        1     28
FT   CHAIN        29    459       VICILIN.
```

```
FT   CONFLICT    375    375       V -> L (IN REF. 2).
FT   CONFLICT    409    409       I -> V (IN REF. 2).
SQ   SEQUENCE   459 AA;  52231 MW;  E2FA390B95451D29 CRC64;

Query Match     21.0%;  Score 745.5;  DB 1;  Length 459;
Best Local Similarity   36.9%;  Pred. No. 5.3e-39;
Matches  159;  Conservative  81;  Mismatches 156;  Indels 35;  Gaps  5;

Qy    250 NPYIYFDERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLPTHLDA 309
Db     34 NPFIFKSNKFQTLFENENGHIRLLQKFDQRSKIFENLQNYRLLEYKSKPHTIFLPQHTDA 93

Qy    310 DAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLLNRDNNERLHIAKFLQTI 369
Db     94 DYILVLSGKAILTVLKPDDRNSFNLERGDTIKLPAGTIAYLVNRDDNEELRVLDLAIPV 153

Qy    370 STPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVF------------- 413
Db    154 NRPGQLQSFLLSGNQNQONYLSGFSKNILEASFNTDYEEIEKVLLEHEKETQHRRSLKD 213

Qy    414 --GQOREGVIIRASQEQIRELTRDDSESRHWHIRGGESSRGPYNLFNKRPLYSNKYGQ 470
Db    214 KRQQSQEENVIVKLSRGQIEELSKNAKSTS----KKSVSSESEPFNLRSRGPIYSNEFGK 269

Qy    471 AYEVKPEDYIRQLQDMDLSVFIANVTQGSMMGPFNTRSTKVVVVASGEADVEMACPHLSG 530
Db    270 FEFITPEKNPQLQDLDIFVNSVEIKESSLLPHYNSRAIVIVTVNEGKGDFLVGQRNEN 329

Qy    531 RHGGRGGGKRHEEEEDVHYEQV------RARLSKREAIVLAGHPVVFVSSGNENLLLF 583
Db    330 QQEQR---KEDDEEEEQGEEEINKQVQONYKAKLSSGDVFVIPACHPVAVKASSNLDLL-- 384

Qy    584 AFGINAQNNHENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESFNSQDQSIFFPGPRQH 643
Db    385 GFGINAENNQRNFLAGDEDNVISQIORVKELAFPGSAQEVDRILENQKQSHFADAQPQQ 444

Qy    644 QQQSPRSTKQQ 654
Db    445 RERGSRETRDR 455

RESULT 11
GLCB_SOYBN
ID   GLCB_SOYBN      STANDARD;      PRT;   439 AA.
AC   P25974;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   BETA-CONGLYCININ, BETA CHAIN PRECURSOR.
GN   CG-4.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93005638; PubMed=2562562;
RA   Harada J.J.; Barker S.J.; Goldberg R.B.;
RT   "Soybean beta-conglycinin genes are clustered in several DNA regions
RT   and are regulated by transcriptional and posttranscriptional
RT   processes.";
RL   Plant Cell 1:415-425(1989).
CC   -!- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC       DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC       CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC   -!- SUBUNIT: THE ALPHA'-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC       VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC   -!- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
CC       VACUOLAR PROTEIN BODIES.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; S44893; AAB23463.1; -.
DR   PIR; JQ0969; FWSYCB.
DR   HSSP; P02853; 2PHL.
DR   INTERPRO; IPR001113; .
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal; Glycoprotein; Multigene family.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   PROPEP       24     25       POTENTIAL.
FT   CHAIN        26    439       BETA-CONGLYCININ, BETA CHAIN.
FT   CARBOHYD    351    351       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   439 AA;  50552 MW;  DBD8AA2A0776088B CRC64;

Query Match     20.8%;  Score 738;  DB 1;  Length 439;
Best Local Similarity   37.1%;  Pred. No. 1.5e-38;
Matches  159;  Conservative  92;  Mismatches 144;  Indels 34;  Gaps  8;

Qy    246 EQSDNPYIF-DERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLP 304
Db     28 EDENNPFYFRSSNSFQTLFENQNVRIRLLQRFNKRSPQLENLRDYRIVQFQSKPNTILLP 87

Qy    305 THLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLLNRDNNERLHIAK 364
Db     88 HHADADFLLFVLSGRAILTVLNNDDRDSYNLHPGDAQRIPAGTTYLVNPHDHQNLKIIK 147

Qy    365 FLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKL-RGVFGQ----QRE 418
Db    148 LAIPVNKPGRYDFFLSSTYQAQQSYLQGFSHNILETSFHSEFEEINRVLFGEEEORQOE 207

Qy    419 GVIIRASQEQIRELTRDDSESRHWHIRGGESSRGPYNLFNKRPLYSNKYGQAYEVKPED 478
Db    208 GVIVELSKEQIRQLSRAKSSS----RKTISSEDEPFNLRSRNPIYSNNFGKFFEITPEK 263

Qy    479 YRQLQDMDLSVFIANVTQGSMMGPFENTRSTKVVVVASGEADVEMACPHLSGRHGGRGGG 538
Db    264 NPQLRDLDIFLSSVDINEGALLLPHFNSKAIVLVINEGDANIELV---------G 310

Qy    539 -----KRHEEEEDVHYEQVRARLSKREAIVLAGHPVVFVSSGNENLLLFAFGINAQNN 592
Db    311 IKEQQOKQKOEEEPLEVQRYRAELSEDDVFVIPAAYP--FVVNATSNLNFLAGINAENN 368

Qy    593 HENFLAGRERNVLQQIEPQAMELAFAAPRKEVEESFNSQDQSIFFPGPRQHQQOSPRSTK 652
Db    369 QRNFLAGEKDNVRQIERQVQELAFPGSAQDVERLLKKQRESYFVDA--QPQOKEEGSKG 426

Qy    653 QQQPLVSIL 661
Db    427 RKGPPSSIL 435

RESULT 12
VCLB_PEA
ID   VCLB_PEA        STANDARD;      PRT;   410 AA.
AC   P02854;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   PROVICILIN PRECURSOR (TYPE B) (FRAGMENT).
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Pisum.
RN   [1]
RP   SEQUENCE FROM N.A. (CLONES PDUB7 AND PDUB4).
RC   STRAIN=CV. FELTHAM FIRST;
```

```
RX   MEDLINE=83220791; PubMed=6687941;
RA   Lycett G.W., Delauney A.J., Gatehouse J.A., Gilroy J., Croy R.R.D.,
RA   Boulter D.;
RT   "The vicilin gene family of pea (Pisum sativum L.): a complete cDNA
RT   coding sequence for preprovicilin.";
RL   Nucleic Acids Res. 11:2367-2380(1983).
-!- FUNCTION: SEED STORAGE PROTEIN.
-!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
-!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
DR   PIR; A03344; FWPMVB.
     HSSP; P50477; 1CAW.
DR   INTERPRO; IPR001113; -.
     PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Glycoprotein; Multigene family; Signal.
FT   SIGNAL        1     15
FT   CHAIN        16   >410        PROVICILIN.
FT   SITE        221    222        CLEAVAGE (POTENTIAL).
FT   CARBOHYD    359    359        N-LINKED (GLCNAC...) (POTENTIAL).
FT   NON_TER     410    410
SQ   SEQUENCE    410 AA;   46385 MW;   8AF68CE85A316FA2 CRC64;

Query Match     19.7%;   Score 699.5;  DB 1;  Length 410;
Best Local Similarity   36.9%;  Pred. No. 3.1e-36;
Matches  147;  Conservative   79;  Mismatches  143;  Indels   29;  Gaps    5;

Qy  249  DNPYFDERSLSTPRFRTEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLPTHLD 308
         ::||| : |||:||   ||| |||:|:||:||::||| ||:| :: |||||||||||||
Db   20  ENPFIFKSNRFQTLYENENGHIRLLQKFDKSRIFENLQNYRLLEYKSKPHTLFLPQYTD 79

Qy  309  ADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLINRDNNERLHIAKFLQT 368
         ||| || |:|||| :::  ||||:|:|:||||||||||||||:|: |||||:||||:||:
Db   80  ADFILVLSGKAILTVLKSNDRNSFNLERGDAIKLPAGSIAYFANRDDNEEPRVLDLAIP 139

Qy  369  ISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVFGQQR---------- 417
         |||||||||||||||:|||||||||||||| ||  ||:||| ::|||:|
Db  140  VNKPGQLQSFLLSGTQNQKSSLSGFSKNILEAAFNTNYEEIEKVLLEQQEQEPQHRRSLK 199

Qy  418  -------EGVIIRASQEQIRELTRDDSESRHWHIRRGGESSRGPYNLFNKRPLYSNKYG 469
                |:|::  ||| |:: ||||::|::  :|::| ||::||:|:::|||||||
Db  200  DRRQEINEENVIVKSRDQIEELSKNAKSSS---KKSVSSESGPFNLRSRNPIYSNKFG 255

Qy  470  QAYEVKPEDYRQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMACPHLS 529
         |:||||||||||||:|||||| |:|: |||||:|:||:|||:   ||||||:|| |||:
Db  256  KFEITPEKNQQLQQDLDIFVNSVDIKVGSLLLPNYNSRAIVIVTVTEGKGDFELVGQR-- 313

Qy  530  GRHGGRGGGKRHEEEEDY--HYEQVRARLSKREAIVVLAGHPVVFVSGGNENLLFAFGI  587
         : |:: :||:| :|||   ||::|  | :|||||||||||||||||:|||||||| |:|
Db  314  NENQGKENDKEEEQQEEETSKQVQLYRAKLSPGDVFVIPAGHPVAINASSDLNLI--GLGI 371

Qy  588  NAQNNHENFLAGRERNVLQQIEPQQAMELAFAFAPRKEVE 625
         ||::||||||||||||:||| ||:| |||:|  :|| |
Db  372  NAENNERNFLAGEEDNVISQVERPVKELAPPGSSHEVD 409
```

```
RESULT  13

ID   VCL_VICFA     STANDARD;     PRT;   463 AA.
AC   P08438;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-JAN-1990 (Rel. 13, Last annotation update)
DE   VICILIN PRECURSOR.
OS   Vicia faba (Broad bean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Vicia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. MINOR;
RX   MEDLINE=88096511; PubMed=3697075;
RA   Weschke W., Baeumlein H., Wobus U.;
RT   "Nucleotide sequence of a field bean (Vicia faba L.var.minor) vicilin
RT   gene.";
RL   Nucleic Acids Res. 15:10065-10065(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. FRIBO;
RX   MEDLINE=88067789; PubMed=3684610;
RA   Bassuener R., van Nong H., Jung R., Saalbach G., Muentz K.;
RT   "The primary structure of the predominating vicilin storage protein
RT   subunit from field bean seeds (Vicia faba L. var. minor cv. Fribo).";
RL   Nucleic Acids Res. 15:9609-9609(1987).
-!- FUNCTION: SEED STORAGE PROTEIN.
-!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
-!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
---------------------------------------------------------------------------
DR   EMBL; Y00506; CAA68559.1; -.
DR   EMBL; Y00462; CAA68525.1; -.
DR   PIR; S06309; S06309.
DR   PIR; A27288; CAA27288.
DR   PIR; A27288; A27288.
     HSSP; P50477; 1CAW.
     INTERPRO; IPR001113; -.
     PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    463        VICILIN.
FT   CONFLICT    257    257        R -> K (IN REF. 2).
FT   CONFLICT    443    443        R -> Q (IN REF. 2).
SQ   SEQUENCE    463 AA;   52694 MW;   5CD09EFEOD5E6B5 CRC64;

Query Match     19.7%;   Score 697.5;  DB 1;  Length 463;
Best Local Similarity   35.1%;  Pred. No. 4.8e-36;
Matches  154;  Conservative   89;  Mismatches  136;  Indels   37;  Gaps    6;

Qy  249  DNPYFDERSLSTPRFRTEGHISVLENFYGRSKLLRALKNYRLVLLEANPNAFVLPTHLD 308
         |||: |: : ||| :||||| |||||:|:||:||::||| ||:|: :::||||||||||
Db   33  DNPFVFESNRFQTLFENENGHIRLLQKFDQHSKLLENLQNYRLLEYKSKPHTIFLPQQTD 92

Qy  309  ADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYLINRDNNERLHIAKFLQT 368
         ||:|| |:||:|| :||:|||||:|:||:|| ::||| |  :::||||::|: :|:|:||
Db   93  ADFILVLLSCKAILTVLLPNDRNSFLERGDTIKLPAGTIGYLVNRDDEEDLRVLDLVIP 152

Qy  369  ISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKLRGVFGQ----------- 415
         :: ||  :|||||:|:||: |||:|||:|||| :||: ||: ||:
Db  153  VNRPGEPQSFLLSGNQNQPSILSGFSKNILEASFNTDYKEIEKVLLEEHGKEKYHRRGLK 212

Qy  416  -----QREGVIIRASQEQIRELTRDDSESRHWHIRRGGESSRGPYNLFNKRPLYSNKYG 469
              ::| |::|::||:|||||:::|::  :|::| ||::||:|:::|||||:||||:
Db  213  DRRQRQEENVIKISRKQIEELKNKASSS---KKSTSSESEPFNLRSREPIYSNKFG 268

Qy  470  QAYEVKPEDYRQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMACPHLS 529
         :::|:||:||:|||:|::||| |  |:||||||:|:||:|||:  |:||||:|| |||:
Db  269  KFEELTPKRNPQLQDLNIFVNYVEINEGSLLLPHYNSRAIVIVTVNEGKGDFELVGQRNE 328

Qy  530  GRHGGRGGGKRHEEEEDVHYEQVR-----ARLSKREAIVVLAGHPVVFVSGGNENLL 582
         :|:: :|||:| :||  ||:::|     :||||:||||||||||||||||||||||||
Db  329  NQQGLR---EEYDEEKEQGEEIRKQVQNYKAKLSPGDVLVIPAGYPVAIKASSNLNLV- 384

Qy  583  FAFGINAQNNHENFLAGRERNVLQQIEPQQAMELAFAAPRKEVEESFNSQDQQSIFPGPRQ 642
         |:||||::||| ::|||| ||:||| |:|: |||:|::||:|| ::||::||:||  ||:
Db  385  -GFGINAENNQRYFLAGEEDNVISQIHKPVKELAPPGSAQEVDTLLENQKQSHFANAQPR 443
```

Left column (RESULT 14)

```
Qy    643 HQQQSPRSTKQQQPLVSIL 661
          ::: ||  | |||
Db    444 ERERGSQEIKDH--LYSIL 460
```

```
RESULT 14
ID   CANA_CANEN      STANDARD;      PRT;    445 AA.
AC   P50477;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CANAVALIN PRECURSOR.
OS   Canavalia ensiformis (Jack bean) (Horse bean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Canavalia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COTYLEDON;
RX   MEDLINE=94143475; PubMed=8310055;
RA   Ng J.D., Ko T.-P., McPherson A.;
RT   "Cloning, expression, and crystallization of canavalin (Canavalia
RT   ensiformis) canavalin.";
RL   Plant Physiol.101:713-728(1993).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE=94143476; PubMed=8310056;
RA   Ko T.-P., Ng J.D., McPherson A.;
RT   "The three-dimensional structure of canavalin from jack bean
RT   (Canavalia ensiformis).";
RL   Plant Physiol. 101:729-744(1993).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; X59467; CAA42075.1; -.
DR   PDB; ICAU; 31-OCT-93; -.
DR   PDB; ICAV; 31-OCT-93.
DR   PDB; ICAW; 31-OCT-93.
DR   PDB; ICAX; 31-OCT-93.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal; 3D-structure.
FT   SIGNAL        1     26
FT   CHAIN        27    445       CANAVALIN.
SQ   SEQUENCE    445 AA;  50326 MW;  3083C5F83A1E9B7 CRC64;
```

```
Query Match  19.0%;   Score 674;  DB 1;  Length 445;  Pred. No. 1.3e-34;
Best Local Similarity  33.3%;
Matches  143;  Conservative  101;  Mismatches  159;  Indels  26;  Gaps  6;
```

```
Qy    235 GGSGRYEEGEEEQS---DNPYIFDERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRL 291
          ||  |  |  |     ||||  :|||  :|  :|:|| |  ::|: | |:||
Db     29 GHSGGEAEDESEESRAQONNPYLFRSNKFPTLFPKNQHGSLRLLQRFNEDTEKLENLRDYRV 88
```

Right column (RESULT 15)

```
Qy    292 VLLEANPNAFVLPFHLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTTFYL 351
          | ||| |:|| | |:||  |:|| ||| |: |||||||| || ||: |:| || || ||
Db     89 LEYCSKPNTLLLPIHSDSDLLVLVLEGQAILVLVLVNPDGRDTYKLDQGDAIKIQAGTPFYL 148
```

```
Qy    352 INRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKL-R 410
Db    149 INPDNNQNRILKFAITTERRPGTVEDFFLSSTKRLPSYLSAFSKNFLRASYDSPYDEIEQ 208
```

```
Qy    411 GVFGQQREGVIIRASQEQIRELTRDDSESRHWHIRRGGESSRGPYNLFNKRPLYSNKYGQ 470
Db    209 TLLQEQEGVIVKMPKDQIOEISKHAQSSS----RKTLSSQDRPFNLRSRDPIYSNNYGK 264
```

```
Qy    471 AYEVKPEDYRQLQDMDLSVFIANYTQGSMMGPFFNTRSTKVVVVASGEADVEMACPHLSG 530
Db    265 LYEITPEKNSQLRDLDILLNCLQMNEGALFVPHYNSRATVILVANEGRAEVELV------ 318
```

```
Qy    531 RHGGRGGGKRHEEEEDVHYEQVR---ARLSKREAIVVLAGHPVVFVSSGNENLLLFAFGI 587
Db    319 ------GLEQQOQOGLESMQLRRYAATLSEGDIIVIPSSFPVALKAASDLNMV--GIGV 369
```

```
Qy    588 NAQNNHIENFLAGRERNVLQOIEPQAMELAFAAPRKEVEESFNSQDQSIFPGPRQHQOQS 647
Db    370 NAENNERNFLAGHKENVIRQIPRQVSDLTFPGSGEEVEELLENOKESYFVDQGQPRHIDAG 429
```

```
Qy    648 PRSTKQQQP 656
          ||: |
Db    430 GKARRAHLP 438
```

```
RESULT 15
ID   CANA_CANGL      STANDARD;      PRT;    445 AA.
AC   P10562;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   CANAVALIN PRECURSOR.
OS   Canavalia gladiata (Sword bean) (Japanese Jack bean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Canavalia.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 27-40.
RC   TISSUE=SEED;
RX   MEDLINE=89296493; PubMed=2740227;
RA   Takei Y., Yamauchi D., Minamikawa T.;
RT   "Nucleotide sequence of the canavalin gene from Canavalia gladiata
RT   seeds.";
RL   Nucleic Acids Res. 17:4381-4381(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEED;
RX   MEDLINE=88111636; PubMed=3338449;
RA   Yamauchi D., Nakamura K., Asahi T., Minamikawa T.;
RT   "cDNAs for canavalin and concanavalin A from Canavalia gladiata
RT   seeds. Nucleotide sequence of cDNA for canavalin and RNA blot
RT   analysis of canavalin and concanavalin A mRNAs in developing seeds.";
RL   Eur. J. Biochem. 170:515-520(1988).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC       CONVICILIN, CONGLYCININ, ETC.).
```

```
DR   EMBL; X06733; CAA29910.1; -.
```

```
DR   EMBL; X15076; CAA33172.1; -.
DR   PIR; S00281; S00281.
DR   HSSP; P50477; 1CAW.
DR   INTERPRO; IPR001113; -.
DR   PFAM; PF00546; Seedstore_7s; 1.
KW   Seed storage protein; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27    445               CANAVALIN.
FT   CONFLICT    161    161               N -> K (IN REF. 2).
SQ   SEQUENCE    445 AA;   50289 MW;  739B59CC2292EA80 CRC64;

Query Match             18.8%;  Score 667;  DB 1;  Length 445;
Best Local Similarity   32.9%;  Pred. No. 3.4e-34;
Matches 141;  Conservative 102;  Mismatches 160;  Indels  26;  Gaps  6;

Qy    235 GGSGRYEGEEEQS---DNPYYFDERSLSTRFRTEEGHISVLENFYGRSKLLRALKNYRL 291
          || | |  :|:|      :| |  :|| : :::: |:|  :| |: ||  | | ::||:
Db     29 GHSGGEAEDESESRAQNNPYLFRSNKFLTLFKNQHGSLRLLQRFNEDTEKLENLRDYRV  88

Qy    292 VLLEANPNAFVLPTHLDADAILLVIGGRGALKMIHHDNRESYNLECGDVIRIPAGTFYL 351
          : || :  ::||:||   |||| :|| | ||   ||  || |  |  :| :  |  ||
Db     89 LEYCSKPNTLLLPHHSDSDLLVLVLEGQAILVLVNPDCRDTYKLDQGDAIKIQAGTPYL 148

Qy    352 INRDNNERLHIAKFLQTISTPGQYKEFFPAGGQNPEPYLSTFSKEILEAALNTQTEKL-R 410
          || ||| | :|  |  |:|    |   |      | |  |||  |  :| |  |:   : |
Db    149 INPDNNQNLRILNFAITFRRPGTVEDFPLSSTKRLPSYLSAFSKNFLEASYDSPYDEIEQ 208

Qy    411 GVFGQQREGVIIRASQEQIRELTRDDSESRHWHIRRGGESSRGPYNLFNKRPLYSNKYGQ 470
          :  :|| ::::|:|:       :: |  |   | |  |:  |||  |::|  |::||||| |
Db    209 TLLQEEQEGVIVKMPKDQIQEISKHIAQSSS---RKTLSSQDKPFNLRSRDPIYSNNYGK 264

Qy    471 AYEVKPEDYRQLQDMDLSVFIANVTQGSMMGPFFNTRSTKVVVVASGEADVEMACPHLSG 530
          | || :  | | :|   | :|      :|: :||:    |  ||| |    | |  |  | |
Db    265 LYEITPEKNSQLRDLDLLNCLQMNEGALFVHYNSRATVILVANEGRAEVELV----- 318

Qy    531 RHGGRGGGKRHEEEEDVHYEQVR----ARLSKREAIVVLAGHPVVFVSSGNENILLFAFGI 587
          : :::  : :::  :  :| :        :::| : : ::  :  | :||  |||::|:||
Db    319 ------GLEQQOQOGLESMQLRRYAATLSEGDILVIPSSFPVALKAASDLNMV--GIGV 369

Qy    588 NAQNNHENFLAGRERNVLQOIEPQOAMELAFAPRKEVEESFNSQDQSIFFPGPRQHQOOS 647
          || | || |:| |:||||:|| : | |::|  : ::|:|:| : ||::| | || || :
Db    370 NAENNERNFLAGNKENVIRQIPRQVSDLTFPGSGEEVELLENQKESYFVDGQPRHIDAG 429

Qy    648 PRSTKQOOP 656
          |:  | :
Db    430 GKARRAHLP 438
```